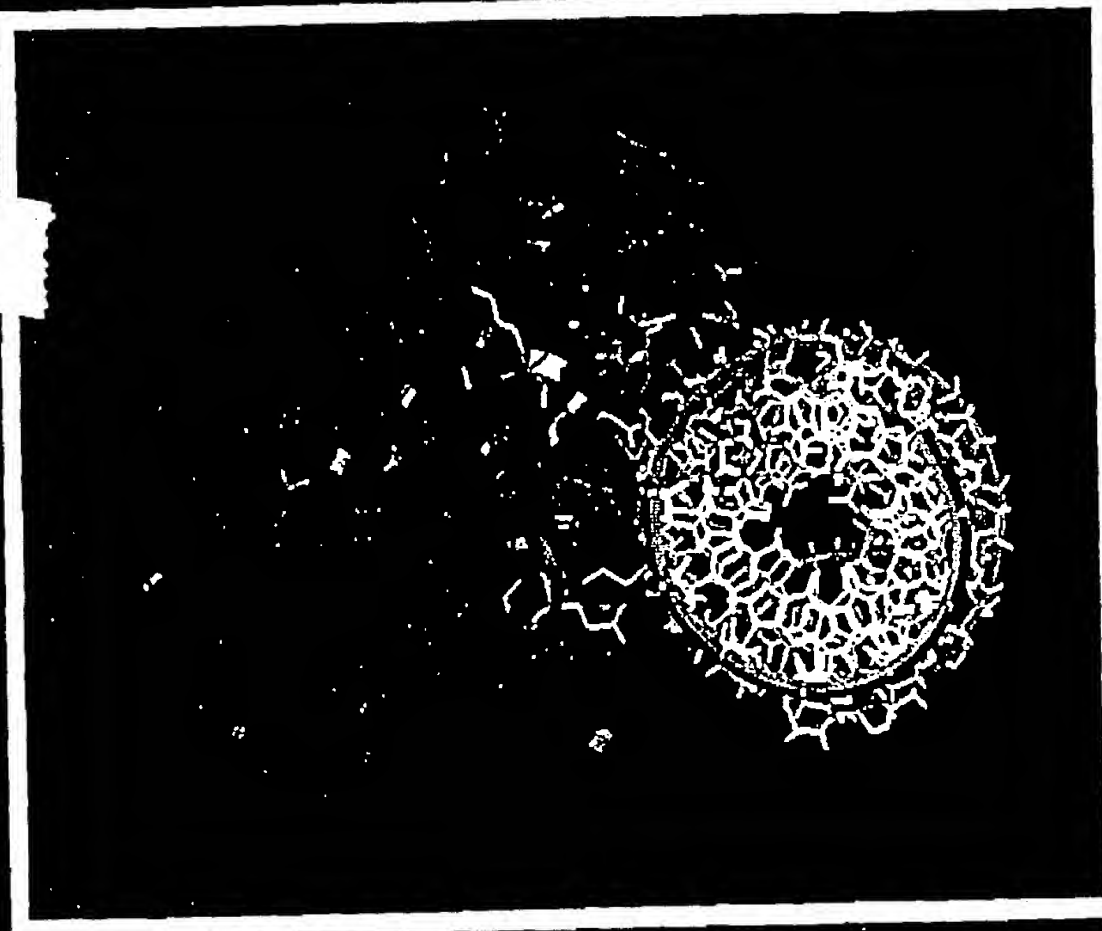


BIOCHEMICAL JOURNAL

Review: The denaturation and degradation
of stable enzymes at high temperatures
by B. M. Daniel, M. Dines and J. H. Drenth

0954-6820/96 \$10.00



Visit the Biochemical Journal on the web at
<http://bmj.oxfordjournals.org>

The Biochemical Society, London © 1996

BJOAK 317(1) 1-327
1996
ISSN 0264-6021

Volume 317
part 1
1 July 1996



Molecular cloning of MADM: a catalytically active mammalian disintegrin–metalloprotease expressed in various cell types

Linda HOWARD*, Xiaohong LU, Sally MITCHELL, Susan GRIFFITHS† and Paul GLYNN‡

Medical Research Council Toxicology Unit, Hodgkin Building, University of Leicester, Lancaster Road, Leicester LE1 9HN, U.K.

A peptide sequence of a metalloprotease purified from bovine brain [Chantry, Gregson and Glynn (1989) *J. Biol. Chem.* **264**, 21603–21607] was used to design an oligonucleotide probe for screening a bovine brain cDNA library. A contig of the two overlapping cDNA clones that were isolated encoded a 748-amino-acid polypeptide with similarity to the disintegrin–metalloprotease precursor proteins of haemorrhagic snake venom. The bovine protein has been named MADM, for mammalian disintegrin–metalloprotease. The predicted mature protein has 534 amino acids arrayed as extracellular metalloprotease and disintegrin (potential integrin-binding) domains, a transmembrane helix and a basic/proline-rich cytoplasmic C-

terminus. Highly conserved homologues of bovine MADM were found in cDNA libraries of rat brain and a human U937 histiocytic lymphoma cell line. A wide variety of mammalian cell lines expressed low levels of MADM mRNA (4.5 and 3.2 kb transcripts) and mature polypeptide (M_r 62000), as assessed by Northern analysis and Western blotting with an antiserum raised to a peptide within the disintegrin domain. MADM appears to be a rather distantly related member of the reprotolysin protein family, which includes both the snake venom disintegrin–metalloproteases and a number of predicted cell-surface disintegrin-containing mammalian proteins.

INTRODUCTION

Most mammalian metalloproteases appear to be involved in extracellular protein catabolism and fall into two broad functional groups. The matrix-degrading metalloproteases (collagenases, gelatinases and stromelysins) are secreted into the extracellular space as latent pro-forms, where they are activated by limited proteolysis. Matrixin activity is tightly controlled at the level of synthesis, which is normally very low and is induced by cytokines and other agents, and by stoichiometric complexing with specific extracellular proteinaceous inhibitors [1,2]. A second group of mammalian metalloproteases are integral membrane proteins and include endopeptidase 24.11 (E24.11) [3], endothelin-converting enzyme (ECE) [4] and meprin [5]. E24.11 and ECE have active-site similarity to the prokaryote enzyme thermolysin, and meprin is a member of the astacin family of metalloproteases [6]. The substrates for E24.11 and ECE are relatively small peptides, but meprin, *in vitro* at least, can act on larger proteins and recently has been reported to degrade extracellular-matrix components [7]. There is great interest in metalloproteases of both these major groups because defects in extracellular peptide/protein catabolism are important in a variety of diseases and these enzymes constitute accessible targets for therapeutic intervention with suitable inhibitors.

Some years ago we isolated from bovine brain myelin membrane preparations a metalloprotease that appeared distinct from either the matrixins or then-defined integral membrane metalloproteases [8]. This protease was subsequently shown to be

present at similar low levels in many bovine tissues [9]. The metalloprotease was identified originally through serendipitous observation of an artefact *in vitro*; i.e. it was detected as the major activity degrading myelin basic protein that had dissociated from its binding sites in myelin membrane preparations [10]. However, as myelin basic protein is a myelin-specific cytoplasmic protein it is unlikely to be the physiological substrate for this widely distributed protease. The aim of this study was to determine the primary structure of the enzyme via cDNA cloning to allow comparison with better-characterized metalloproteases.

EXPERIMENTAL

Protein biochemistry

Metalloprotease was isolated from bovine brain myelin membrane preparations as described previously [8]. N-terminal sequences of the purified metalloprotease and of seven tryptic peptides were determined by automated Edmann degradation by Dr. A. Willis (MRC Immunochimistry Unit, University of Oxford, Oxford, U.K.).

cDNA cloning

Standard procedures detailed by Sambrook et al. [11] were used throughout. Sequence (TTADEKDPTNPF) from one tryptic peptide was used to design a synthetic oligonucleotide probe ACIACIGCIGATGAGAAGGACCCCIACIAACCCITT, where

Abbreviations used: E24.11, endopeptidase 24.11; ECE, endothelin-converting enzyme; MADM, mammalian disintegrin–metalloprotease; UTS, untranslated sequence.

* Present address: Department of Cell Biology and Lombardi Cancer Center, Georgetown University Medical Center, Washington, DC 20007, U.S.A.

† Present address: Centre for Medicines Research, Carshalton, Surrey SM5 4DS, U.K.

‡ To whom correspondence should be addressed.

The nucleotide sequences reported in this paper have been submitted to GenBank/EMBL Data Bank with the accession numbers Z21961 (bovine), Z48444 (rat) and Z48579 (human).

GCGCGCGGACGAGAGATGGTCTGCTGAGATGTTAATCTGCTCTCTGCGGTGCGGGGCTGGAGGTGAGTATGGAAATCCTTTAAATAATACATTAGACATTATGAA- 115
 MetValLeuLeuAspValLeuLeuLeuSerTrpValAlaGlyLeuGlyGlyGlnTyrGlyAsnProLeuAsnLysTyrIleArgHisTyrGlu- 133
 GATTATCATATGATGATGATTACACAAAAACAGCTAGCGTGGCAAGAGCAGTATCACATGAGGACAGTCTTTACGGCTAGATTTCATGCTCATGGAAGACATTCAACCTT- 235
 GlyLeuSerTyrAspValAspSerLeuHisGlnLysHisGlnArgAlaLysArgAlaValSerHisGluAspGlnPheLeuArgLeuAspPheHisAlaHisGlyArgHisPheAsnLeu- 73
 CGAATGAAGAGGATACCTCTCTTTTCACTGAGGAGTTAGCGTGGAAACATCAATGCAGTACTGATTATGATCTCTCATATTTACACTGGACATATTTATGGTGAAGACGAGT- 355
 ArgMetLysArgAspThrSerLeuPheSerGluGluPheArgValGluThrSerAsnAlaValLeuAspTyrAspThrSerHisIleTyrThrGlyHisIleTyrGlyGluGluGlySer- 113
 TTAGCATGGTCTGTTATTGATGAAGATTGAAGGATTCACTCAGACTCATGGTGGCAGGATTTATGTTGAACAGCAGAGAGATATTAAGACCGAACTCTGCCATTTCACTCT- 475
 LeuAlaMetGlyLeuLeuLeuMetGluAspLeuLysAspSerPheArgLeuMetValAlaArgPheTyrValGluProAlaGluArgTyrIleLysAspArgThrLeuProPheHisSer- 153
 GTCATTTATCATGAAGATGATATTAAGTATCCCCATAATATGGTCCACAGGGCGTGTGCAGATCATTCACTGTTTGAAGAATGAGGAAGTACCAGATGACTGGTGTAGAAGAGTA- 595
 ValIleTyrHisGluAspAspIleLysTyrProHisLysTyrGlyProGlnGlyArgCysAlaAspHisSerValPheGluArgMetArgLysTyrGlnMetThrGlyValGluGluVal- 193
 ACACAGACACCTCAAGAAAAACATGCTATTATGGTCCAGAACTCTGAGGAAAAACGTAGAACTGTAGCTGAAAAAATACCTGTGAGCTTTATATTACAGCCGATCATCTGTTCTT- 715
 ThrGlnThrProGlnGluLysHisAlaIleAsnGlyProGluLeuLeuArgLysLysArgThrThrValAlaGluLysAsnThrCysGlnLeuTyrIleGlnThrAspHisLeuPhePhe- 233
 AAATATTACGAACAGAGAGAGCTGTGATTGCCAGATATCCAGTCATGTTAAAGCAATTGACACAATTTACAGACAACAGACTTCTCGGAATCCGTAACATCAGTTTCATGGTGA- 835
 LysTyrTyrGlyThrArgGluAlaValIleAlaGlnIleSerSerHisValLysAlaIleAspThrIleTyrGlnThrThrAspPheSerGlyIleArgAsnIleSerPheMetValLys- 273
 CGCATAAGAATCAACAACTGCTGATGAGAGGACCTACAAATCCATTCCGTTTTCCAAATATTGGTGTGGAGAGTTTCTGGAGCTGAATCTGAGCAGAACTCATGATGACTACTGT- 955
 ArgIleArgIleAsnThrThrAlaAspGluLysAspProThrAsnProPheArgPheProAsnIleGlyValGluLysPheLeuGluLeuAsnSerGluGlnAsnHisAspAspTyrCys- 313
 TTGGCGTACGTTTTACAGATGAGATTGTGATGATGGTCTCTGGTCTGGCGTGGTGGAGCACCTTCAGGAAGCTCTGGAGGAATATGTGAAAAAGTAAGCTCTATTGAGATGGT- 1075
 LeuAlaTyrValPheThrAspArgAspPheAspAspGlyValLeuGlyLeuAlaTrpValGlyAlaProSerGlySerSerGlyGlyIleCysGluLysSerLysLeuTyrSerAspGly- 353
 AAGAGAAGTCTTTAAACACTGGAATTATTACTGTTCAAGTATGGTCTCAGTACCCGCAAGTCTCTCAGTATACGTTTGTCTATGAAGTGGACATAACTTTGGATCTCCGAT- 1195
 LysLysLysSerLeuAsnThrGlyIleIleThrValGlnAsnTyrGlySerHisValProProLysValSerHisIleThrPheAlaHisGluValGlyHisAsnPheGlySerProHis- 193
 GATTCTGGAACAGAGTGCCTCCAGGAGATCTAAGATTAGGACAAAAAGAAATGGCACTACATCATGTATGCAAGAGCAACATCTGGGACAACTTAACACAAATAAATCTCA- 1315
 AspSerGlyThrGluCysThrProGlyGluSerLysAsnLeuGlyGlnLysGluAsnGlyAsnTyrIleMetTyrAlaArgAlaThrSerGlyAspLysLeuAsnAsnAsnLysPheSer- 433
 CTCTGTAGTATTAGAAATATAAGTCAAGTCTCTGAGAAGAGAGAAACACTGTTTTGTTGAATCTGCGCACTATTGTTGGAATCGGATGGTAGAACAGGTGAAGATGTGATTGT- 1435
 LeuCysSerIleArgAsnIleSerGlnValLeuGluLysLysArgAsnAsnCysPheValGluSerGlyGlnProIleCysGlyAsnGlyMetValGluGlnGlyGluGluCysAspCys- 473
 GGTATAGTGACCGTGTAAAGACAGTGTCTACGATCCAAATCAGCGGAGGAGGAAAAATGCAAGTGAACCTGGAAGCAGTGCAGTCCAGTCAAGTCCCTGTTGTACAGCA- 1555
 GlyTyrSerAspGlnCysLysAspGluCysCysTyrAspAlaAsnGlnProGluGlyLysLysCysLysLeuLysProGlyLysGlnCysSerProSerGlnGlyProCysCysThrAla- 513
 CATTGTGCAATCAAGTCAAAACTGAAAGTGTGGGATGATTCAAGTGTGCAAAAGAGGAATATGTATGGCATCACAGCTCTCTGCCAGCTCTGATCCTAAACCGAACTTCACA- 1675
 HisCysAlaPheLysSerLysThrGluLysCysArgAspAspSerAspCysAlaLysGluGlyIleCysAsnGlyIleThrAlaLeuCysProAlaSerAspProLysProAsnPheThr- 553
 GACTGTAATAGACATACGCAAGTGTGATTAAATGGCAATGTGCAAGTCTCTATCTGAGAAACATGCGCTGGAGGAGTGTACCTGTGCCAGTCTGATGGCAAGATGATAAGCAATTA- 1795
 AspCysAsnArgHisThrGlnValCysIleAsnGlyGlnCysAlaGlySerIleCysGluLysHisGlyLeuGluGluCysThrCysAlaSerSerAspGlyLysAspAspLysGluLeu- 593
 TGCCATGTCTGCTATGAAGAGATGGAGCCATCAACTGTGCGCAGTACAGGTCTGTGCGAGTGAACAGTACTTCCCTGGTGAAGTATCACCTGCAAGCTGGATCCCCATGCAAT- 1915
 CysHisValCysCysMetLysLysMetGluProSerThrCysAlaSerThrGlySerValGlnTrpAsnLysTyrPheLeuGlyArgThrIleThrLeuGlnProGlySerProCysAsn- 633
 GATTTAGAGGCTACTGTGATGTTTTCATGCGGTGCGATTAGTAGTGTGCTGCTAGCGAGGCTTAAAAAGCAATTTTCAGTCCAGAGCTCTATGAAACATAGCTGAATGG- 2035
 AspPheArgGlyTyrCysAspValPheMetArgCysArgLeuValAspAlaAspGlyProLeuAlaArgLeuLysLysAlaIlePheSerProGluLeuTyrGluAsnIleAlaGluIle- 673
 ATTGTGCTTACTGGTGGCAGTATTACTTATGGAAATGCGCTGATCATGTTAATGGTGGTITTTAAGATATGCAAGTGTACACTCCAGTAGTAATCCAAAGTTGCGTCTCTCT- 2155
 IleValAlaTyrTrpTrpAlaValLeuLeuMetGlyIleAlaLeuIleMetLeuMetAlaGlyPheIleLysIleCysSerValHisThrProSerSerAsnProLysLeuProPro- 713
 AAACCTCTCCAGGCACTTTAAAGAGGAGAGACCTCCCGAGCCATCAACAGCCCGAGGTGAGAGGCGCGGAGGTTATCAGATGGACACATGAGAGT- 2260
 LysProLeuProGlyThrLeuLysArgArgArgProProGlnProIleGlnGlnProGlnArgGlnArgProArgGluSerTyrGlnMetGlyHisMetArgArg- 748
 caactcagcttttgccttggtcttctctagtgcctcgaatgggaaccttccctccaaagagaacctattaaat- 2335
 catcatcccccaatgaacctccacacataacagtaga- 2390

Figure 1 For legend see facing page

best guesses were made for the final base in 2-fold redundant codons and inosine was used in cases of 4-fold redundancy. The probe was end-labelled with [32 P]phosphate and used to screen approx. 800 000 plaques of an oligo-(dT)-primed λ gt10 bovine brain cDNA library (a gift from Dr. A. Jackson, University of Cambridge, Cambridge, U.K. [12]). Hybridization and washing were done at 46 °C in 2×0.15 M NaCl/0.015 M sodium citrate (SSC)/0.1% SDS. Positive plaques were subjected to three rounds of screening until a single clone with a 0.9 kb insert was isolated. Insert cDNA was excised with *Eco*RI, purified, random-labelled with [32 P]dCTP and used to rescreen the same brain library with a more stringent final wash ($1 \times$ SSC/0.1% SDS at 65 °C). Positive plaques were taken through three further rounds of screening. A 2.4 kb positive clone was isolated; insert cDNA was excised using *Hind*III and *Bam*HI and then ligated into pUC18. Double-stranded DNA of the recombinant plasmids was sequenced with Sequenase kit version II (USB) using initially λ gt10 forward and reverse primers and subsequently primers designed from the sequences obtained. Each base on both strands was sequenced at least three times, and the data were analysed using PC/GENE programs (Intelligenetics). The initial 0.9 kb clone comprised nt 1–930 of the sequence shown in Figure 1 and was truncated at its 3' end by an internal *Eco*RI site. The 2.4 kb clone was truncated at its 5' end and comprised nt 47–2385 of Figure 1. A contig was constructed from the two partial clones, and similarities between the sequence of the deduced bovine polypeptide and other proteins in the databases were sought using BLAST and TFASTA programmes.

A 1.3 kb portion of the isolated bovine cDNA (corresponding to nt 662–1980 of Figure 1) was amplified by PCR, radiolabelled (Megaprime Kit; Amersham) and used as a probe to screen a λ gt10 rat brain cDNA library (catalogue no. RL1403a; Clontech). Filters were hybridized in the presence of 20% formamide at 42 °C and finally washed in $2 \times$ SSC at 45 °C. A positive clone (designated clone E) was isolated through three rounds of screening; its 2.1 kb insert was subcloned into pBluescript SK, sequenced and shown to have a 5' end equivalent to nt 924 of the bovine cDNA sequence. Further 5' sequence was obtained by the rapid amplification of cDNA ends technique using a kit from Clontech and poly(A)⁺-RNA isolated (see below) from rat brain.

Identical procedures were used to isolate a clone designated pXL from a λ gt10 human histiocytic lymphoma cell (U937) cDNA library (Clontech catalogue no. HL1036a; a gift from Dr. G. Harper, SB Pharmaceuticals). pXL contained a 1.3 kb insert with a 5' end equivalent to nt 1335 of the bovine cDNA sequence. Additional 5' sequence for the human homologue of the bovine metalloprotease was obtained by PCR on aliquots of the U937 cDNA library between a primer homologous to the 5' end of pXL and the λ vector primer.

Northern and Western blotting

Poly(A)⁺-RNA was isolated from human cell lines by centrifugation in caesium chloride gradients [11] followed by chromatography on oligotex-(dT) beads (Qiagen). Aliquots (10 μ g) were run on formamide gels, transferred to Hybond-N+ membranes and hybridized with a radiolabelled human cDNA probe (pXL). The membranes were washed at high stringency (65 °C;

0.1 \times SSC/0.1% SDS) and then exposed to X-ray film with intensifying screens for 2 days at –80 °C.

A peptide (FDANQPEGKKC) corresponding to amino acids 485–495 of the deduced rat and human polypeptide sequences was synthesized (by Professor N. Groome, Oxford Brookes University, Oxford, U.K.), coupled to keyhole limpet haemocyanin with *m*-maleimidobenzoyl-*N*-hydroxysuccinimide ester and used to immunize rabbits. Antisera were preadsorbed with (keyhole limpet haemocyanin)-Sepharose and tested by Western blotting against glycoproteins that were isolated by concanavalin A-Sepharose chromatography from various tissues and cell lines as described previously, except that alkaline phosphatase-labelled Protein A and bromochloroindoyl phosphate/Nitro Blue Tetrazolium substrate were used to reveal reactivity on the blots [9]. Our earlier work with a monoclonal antibody raised to the purified bovine metalloprotease had established that it comprised less than 0.01% of total tissue protein and could not be detected by Western blotting of crude homogenates [9].

RESULTS AND DISCUSSION

General features of the cloned metalloprotease cDNA and domain structure of the encoded protein

A contig of the sequences of the 0.9 kb and 2.4 kb cDNA clones isolated from the bovine brain library (see the Experimental section) and the predicted protein it encodes are shown in Figure 1. The cDNA clone has a 13 nt poly(A)⁺ tail at its 3' end, and 57 nt upstream of this is the hexamer ATTAAA, which is the most frequently observed deviation from the consensus polyadenylation signal AATAAA [13]. The first initiator codon, ATG, begins at nt 17, is flanked by nucleotides (A-14 and G-20) conforming to a Kozak consensus sequence [14] and is followed by an open reading frame encoding a protein of 748 amino acids. All the peptide sequences of the metalloprotease determined by Edman degradation are found within this predicted protein, including the mature N-terminus, which begins at Thr-214.

The mature protein contains four potential glycosylation sites (Asn residues 267, 278, 439 and 551) of which (at least) Asn-278 is actually modified, as indicated by its characteristics during Edman sequencing of the tryptic peptides. The glycosylated metalloprotease purified from bovine brain has an apparent M_r of about 62 000 in reducing SDS/PAGE (results not shown). This value compares reasonably well with a calculated M_r for the predicted mature polypeptide (Thr-214–Arg-748) of 59 200. The predicted protein has a typical eukaryotic signal sequence (residues 1–19) and a potential transmembrane helix (residues 673–696).

A sequence database search revealed significant similarity (18.6% identity over the region of overlap; amino acids 46–653 in Figure 1) between the deduced protein sequences of the bovine protease and jararhagin, a metalloprotease present in the venom of the pit viper, *Bothrops jararaca* [15]. Jararhagin and closely related proteins in other haemorrhagic snake venoms have four domains [16]: 1, pre-prosequence; 2, protease; 3, disintegrin; and 4, cysteine-rich domains. In addition to the four domains in jararhagin, the bovine protein has a potential transmembrane helix and a predicted cytoplasmic C-terminus of 51 residues rich

Figure 1 cDNA and deduced protein sequences for the bovine metalloprotease

The predicted eukaryotic signal sequence is marked with arrowheads (▲▲▲▲▲▲). The mature N-terminus, Thr-214, is marked with an asterisk. Regions of protein sequence determined by Edman degradation are underlined. Solid circle symbols mark potential N-glycosylation sites. The internal *Eco*RI site and polyadenylation signal are bold and underlined. The catalytic zinc-binding site (residues 383–393) is shown in bold uppercase letters. The putative membrane-spanning domain is double underlined and bold. The 3' untranslated nucleotide sequence begins at the stop codon *taa* (nt 2261–2263) and is shown in lower case letters.

BOVINE	MVLLPVLILLSSWAGLGGOYGNPKYIRHYEGLSYDVS LKQKORAK	50
BOVINE	RAVSHEDGFLRLOFHAKGRHFNLRHKDTS LFSFEFRVETSNVLDYDYS	100
BOVINE	HIYTGHIYCEEGSLAMGLLME DLKDSFRLMVAFFYVEPAERYIKDRTLP	150
BOVINE	FHSVIYHEDDINKYPIKYGPGRCADHSVFERMRKYQMTGVEEVTQTPQEK	200
BOVINE RAT	HAINGPELLRKRTTVAEIONTCQLYIOTDHLFFKYGYTRAVIAQISSHV GPELLAKORTILPERNTCQLYIOTDHLFFKYGYTRAVIAQISSHV	250
BOVINE HUMAN RAT	KAITDIYQTTDFSGIRNISTHVKIRINTTADKDPNPFPRFNIGVEKF OTTDFSGIRNISTHVKIRINTTADKDPNPFPRFNISVEKF KAITDAIYQTTDFSGIRNISTHVKIRINTTADKDPNPFPRFNIGVEKF	300
BOVINE HUMAN RAT	LELNSEQHDDYCLAYVETDRDFDDGVLGLAWGAPSGSSGGICEKSKLY LELNSEQHDDYCLAYVETDRDFDDGVLGLAWGAPSGSSGGICEKSKLY LELNSEQHDDYCLAYVETDRDFDDGVLGLAWGAPSGSSGGICEKSKLY	350
BOVINE HUMAN RAT	SDGKKS LNTGII TVQYGSHPVPPKVSHTFAHEVGHNFSGPHDSGTECT SDGKKS LNTGII TVQYGSHPVPPKVSHTFAHEVGHNFSGPHDSGTECT SDGKKS LNTGII TVQYGSHPVPPKVSHTFAHEVGHNFSGPHDSGTECT	400
BOVINE HUMAN RAT	PGESKNLGQKENGNYIHYARATSGDKLNNKFSLSLSIRNISQVLEKORNN PGESKNLGQKENGNYIHYARATSGDKLNNKFSLSLSIRNISQVLEKORNN PGESKNLGQKENGNYIHYARATSGDKLNNKFSLSLSIRNISQVLEKORNN	450
BOVINE HUMAN RAT	CFVESGQPICGNGMVEQGEEDCGYSQCKDECCFDANQPEGRKCKLKP CFVESGQPICGNGMVEQGEEDCGYSQCKDECCFDANQPEGRKCKLKP CFVESGQPICGNGMVEQGEEDCGYSQCKDECCFDANQPEGRKCKLKP	500
BOVINE HUMAN RAT	KOCSPSQGPCTAHCAPKSKTEKRDSDCAKEGICNGITLCPASDPKP KOCSPSQGPCTAHCAPKSKTEKRDSDCAKEGICNGITLCPASDPKP KOCSPSQGPCTAHCAPKSKTEKRDSDCAKEGICNGITLCPASDPKP	550
BOVINE HUMAN RAT	NFTDCNRHTQVCINGOCAGSICEKYLEECTCASSDGKDDKELCHVCCMK NFTDCNRHTQVCINGOCAGSICEKYLEECTCASSDGKDDKELCHVCCMK NFTDCNRHTQVCINGOCAGSICEKYLEECTCASSDGKDDKELCHVCCMK	600
BOVINE HUMAN RAT	KMEPSTCASTGSGVQWNYFLGRTITLQPGSPCNDFRGYCDVFMCRVLDA KMEPSTCASTGSGVQWNYFLGRTITLQPGSPCNDFRGYCDVFMCRVLDA KMEPSTCASTGSGVQWNYFLGRTITLQPGSPCNDFRGYCDVFMCRVLDA	650
BOVINE HUMAN RAT	DGFLARLKKAI FSPELYENIAEWIVAHWVAVLLMGIALIMHAGFIKICS DGFLARLKKAI FSPELYENIAEWIVAHWVAVLLMGIALIMHAGFIKICS DGFLARLKKAI FSPELYENIAEWIVAHWVAVLLMGIALIMHAGFIKICS	700
BOVINE HUMAN RAT	VHTPSSNPKLPPPKPLPGTLKRRRPPQPIQOPQORPPRESYQMGHRR* VHTPSSNPKLPPPKPLPGTLKRRRPPQPIQOPQORPPRESYQMGHRR* VHTPSSNPKLPPPKPLPGTLKRRRPPQPIQOPQORPPRESYQMGHRR*	748

Figure 2 Deduced protein sequences from bovine, rat and human MADM cDNAs

Non-identical residues are marked with solid diamond symbols, and positions of the extended zinc-binding site (HEV...PH) and the predicted transmembrane helix (WIV...GFI) are underlined with stars. Stop codons are indicated by a single asterisk.

in proline (23.5%) and basic amino acids (23.5%). The greatest similarity between jararhagin and the bovine protein was found over the disintegrin (potential integrin-binding) domain (36.6% identity). In recognition of this homology the bovine protein was named MADM, for mammalian disintegrin-metalloprotease.

The long pre-prosequence of MADM (residues 1-213) contains a single cysteine (Cys-173) that, by analogy with metalloproteases of the matrixin and snake venom families, may interact with the zinc ion at the active site of the protease domain and maintain the proenzyme in a latent state [17,18]. The prosequence ends with the tetrabasic motif RKKR, a typical cleavage site for furin-like proteases in the Golgi, and is characteristic of proteins that are transported to the cell surface by a constitutive pathway rather than one that involves storage in granules [19,20]. This predicts that MADM will be presented at the cell surface with the mature N-terminus of its protease domain exposed and its catalytic zinc ion in an active form. This is quite different to the majority of secreted latent proenzymes of the matrixin family but similar to the processing recently described for prostromelysin-3 [21]. The prosequence of a membrane-type

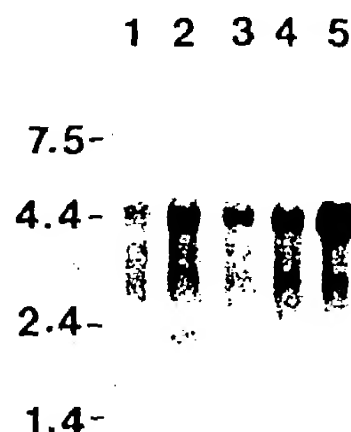


Figure 3 Expression of MADM mRNA in different human cell lines

Poly(A)⁺ RNA was isolated from the cell lines indicated, and equal amounts (10 µg/lane) were subjected to Northern analysis as described in the Experimental section. The migration of standard RNA size markers (kb) is indicated. Lane 1, Y79; 2, U937; 3, MRC-5 SV1; 4, D98/AH2; 5, A549.

matrixin (membrane-type matrix metalloprotease) terminates with the motif RKKR, and thus pro-membrane-type matrix metalloprotease may also be processed by furin [22].

The protease domain of MADM (residues 214-455) contains an extended zinc-binding site, HEVGHNFSGSPH (residues 383-393), and is followed by disintegrin and cysteine-rich domains (residues 456-550 and 551-653) that contain 15 and 13 cysteine residues respectively. The cysteine-rich domain is linked to the predicted transmembrane helix by a stretch of 19 amino acids.

Rat and human homologues of MADM and distribution in tissues and cell lines

Part of the 2.4 kb bovine MADM cDNA clone was used to screen libraries of rat and human cDNA (see the Experimental section). Partial cDNA clones encoding proteins highly similar to bovine MADM were isolated, not only from a rat brain library but also from a human U937 histiocytic lymphoma cell line library, indicating that MADM is not a brain-specific transcript. The 3' untranslated sequences (UTSs) of the rat and human cDNAs were incomplete, as indicated by their lack of a poly(A)⁺ tail, but were still substantially longer than the 3' UTS of the bovine MADM cDNA clone. Nevertheless, the polypeptide sequences deduced from the coding region of these cDNAs were 97% identical over the region of overlap (Figure 2). The human cDNA clone (pXL) was used as a probe for Northern analysis of poly(A)⁺-RNA isolated from U937 cells and four other human cell lines and revealed transcripts of 4.5 and 3.2 kb, the larger being especially prominent in A549 cells (Figure 3). The reason for the presence of two human MADM mRNA species is not clear, but may reflect differences in the lengths of the 5' and 3' UTSs.

We have reported the use of a monoclonal antibody, CG4 which recognizes a species-specific epitope of the bovine brain metalloprotease, to demonstrate its presence at similar low levels in various bovine tissues, including kidney, spleen, lung, adrenal heart and peripheral nerve [9]. To allow screening of human and rodent cells and tissues we raised an antiserum to a disintegrin domain peptide sequence conserved between these two species (see the Experimental section). This antiserum readily detected a band of *M*_r 62000 at similar levels in Western blots of dithiothreitol-reduced glycoproteins isolated from a variety of

rat and bovine tissues (brain, kidney, lung, spleen, ovary and testis; results not shown) and from a diverse range of mammalian cell lines, including bovine kidney (MDBK), mouse erythro-leukaemia (MEL), rat oligodendroglioma (33B), human carcinomas from colon (Colo 201 and HT115) and lung (A549), retinoblastoma (Y79) and histiocytic lymphoma (U937) (Figure 4). Lower levels of the M_r 62000 polypeptide were present in a bovine lymphocyte line (BL-3), human carcinomas from cervix (D98) and larynx (Hep2) and fibroblast lines from normal embryo (S1814.PB5) and simian virus 40-transformed foetal lung (MRC-5 SV1) (Figure 4).

The M_r 75000 positive band in glycoproteins from D98 cervical carcinoma (Figure 4, lane 4) may represent an immature, partially processed form of MADM, although it appears rather small for the predicted full proMADM (approx. M_r 85000). However, as this band was not seen in any other cell types its identity was not pursued further. Preincubation of the antiserum with the immunizing peptide, but not with a control peptide, blocked the staining of the M_r 62000 band on glycoprotein blots and thereby demonstrated the specificity of the reactivity (results not shown).

Our earlier studies [9] on the distribution of MADM in tissue homogenates left open the possibility that the protein was present solely in a common cell type (such as vascular endothelium or connective tissue fibroblasts). The present experiments demonstrate that MADM is expressed at low levels in a wide range of epithelial and haemopoietic cells that are both adherent and non-adherent *in vitro*. Nonetheless, it should be noted that the amount of MADM in glycoprotein fractions from any of the cell lines examined was not greater than that in glycoproteins extracted from whole tissues including brain and kidney (results not shown).

Comparison of MADM with other disintegrin-containing proteins

Since 1992 it has become clear that a number of mammalian proteins and the snake venom disintegrin-metalloproteases are members of the same family, the repolysins [23]. The domain structures of the predicted mature forms of representative mammalian members of this family and of jarrahagin are compared in Figure 5. PH-30 [24] and epididymal apical protein-1 [25] are expressed in the male reproductive tract, and there is good evidence that PH-30 is involved in sperm-egg recognition and fusion [26]. PH-30 exists as an α - β heterodimer in which the potentially catalytically active metalloprotease domain plus most of the disintegrin domain of the α precursor and the catalytically inactive metalloprotease domain of the β precursor are removed during maturation of the protein [27]. Meltrin- α is involved in myoblast fusion and appears to require further processing to a form lacking most of the protease domain to function in this role [28]. The putative functions of the other mammalian members of this family are unknown.

Among the mature proteins depicted in Figure 5, MADM appears most similar to meltrin- α (see above) and MS2 [29]. MS2 was isolated by subtractive hybridization methods from a mouse macrophage cell line and appears to be a macrophage-specific transcript. Northern analysis of macrophage RNA reveals two abundant MS2 transcripts (3.1 and 4.6 kb) whose expression is substantially enhanced by phorbol ester treatment [29]. By contrast, MADM mRNA transcripts (3.2 and 4.5 kb) are found at low levels in a variety of cell types (see Figure 3), and expression (at least in U937, Y79 and A549 cells) is not greatly enhanced by phorbol ester treatment (results not shown). Like MADM, MS2 has a basic/proline-rich cytoplasmic tail, a transmembrane helix and extracellular disintegrin and potentially



Figure 4 Expression of MADM polypeptide in various mammalian cell lines

Glycoproteins were isolated from the cells indicated, and equal amounts (6 μ g/lane) were run on SDS/PAGE, transferred to nitrocellulose blots and probed with rabbit antiserum to a disintegrin domain peptide (see the Experimental section). The migration of standard protein size markers ($10^{-3} \times M_r$) is indicated. Lanes 1-9 are human cell lines and lanes 10-13 are other mammalian cell lines: 1, Y79; 2, U937; 3, MRC-5 SV1; 4, D98/AH2; 5, A549; 6, Hep2; 7, Colo 201; 8, HT115; 9, S1814.PB5; 10, MDBK; 11, BL-3; 12, MEL; 13, 33B.

active protease domains. Nonetheless, MS2 and MADM differ structurally in that mature MS2 (apparent M_r 89000) has an N-terminal extension beyond the protease domain [29].

Within the repolysin family, epididymal apical protein-1, metalloprotease disintegrin/cysteine-rich protein [30,31] and PH-30 β have catalytically inactive metalloprotease domains, since they lack some of the critical active-site amino acids (HEXXHXXGXXH), whereas MADM, MS2, meltrin- α and the PH-30 α precursor have a full complement of these residues (Figure 6a). The third His in this sequence is followed by Asp, a signature for members of the repolysin family as opposed to the astacins and matrixins that have Glu and Ser, respectively, in this position [32].

It has been suggested that a 'methionine turn' following the extended zinc-binding site plays an important role in the structural integrity of the active site of all these metalloprotease families, collectively known as metzincins [33]. In the region between the zinc-binding site and the conserved methionine,

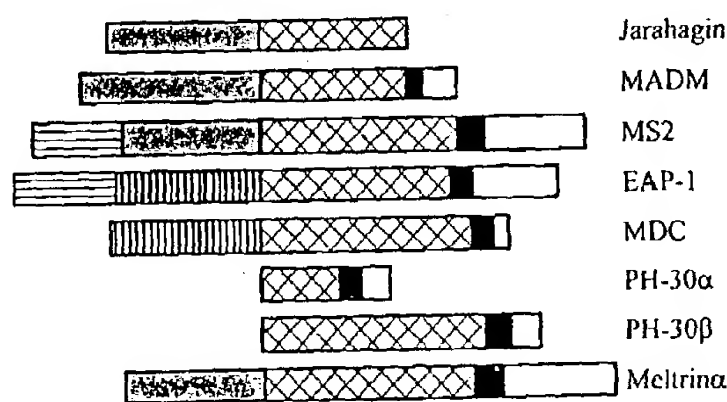
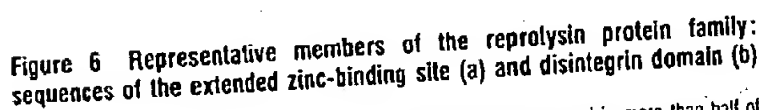


Figure 5 Representative members of the repolysin protein family: domain structure of the mature polypeptides

Mature polypeptides that lack some of the critical residues, HEXXHXXGXXH, required for catalytic site zinc binding are shown with a metalloprotease-like domain (see also Figure 6a). EAP-1, epididymal apical protein-1; cloned from rat and monkey [25]. MDC, metalloprotease-disintegrin/cysteine-rich protein; present in various tissues as two isoforms with (769 amino acids [30]) and without (524 amino acids [31]); a transmembrane domain of which the larger is more abundant and shown in the Figure. [▨], N-terminal domain; [▤], metalloprotease-like domain (catalytically inactive); [▦], disintegrin/cysteine-rich domain; [■], transmembrane helix; [□], cytoplasmic domain.



MADM differs from the other reprolysins in having a greater total number of intervening residues but also lacks two of the three conserved cysteines (Figure 6a).

The 15 conserved cysteines in the disintegrin domains of the mammalian reprotlysins and jararhagin are shown numbered in Figure 6(b). Relative to the other reprotlysins MADM has a six-amino-acid insertion after Cys-6 but lacks: a conserved Gly after Cys-11; one amino acid between Cys-12 and -13; conserved Asp and Glu after Cys-13. Thus from sequence comparison of both the metalloprotease active site and disintegrin domains, MADM appears rather different from the other reprotlysins, and phylogenetic analysis suggests that MADM diverged early from other known members of this family [34].

We thank SmithKline Beecham (SB) Pharmaceuticals and the Multiple Sclerosis Society (U.K.) for financial support, A. Willis for protein sequencing, A. Jackson for the bovine brain cDNA library, N. Groome for peptide synthesis, C. Southan (SB Pharmaceuticals) for help with initial database search, M. Manson (MRC Toxicology Unit) and G. Harper (SB Pharmaceuticals) for support and encouragement.

- 1 Matrisian, L. (1992) *Bioessays* **14**, 455-463
- 2 Nagase, H. and Salvesen, G. (1993) in *Innovations in Proteases and their Inhibitors* (Aviles, F. X., ed.), pp. 315-332. Walter de Gruyter, Berlin
- 3 Roques, B. P., Noble, F., Dauge, V., Fournie-Zouluski, H.-C. and Beaumont, A. (1993) *Pharmacol. Rev.* **45**, 88-146
- 4 Shimada, K., Takahashi, M. and Tanzawa, K. (1994) *J. Biol. Chem.* **269**, 18275-18278
- 5 Jiang, W., Gorbea, C. M., Flannery, A. V., Beynon, R. J., Grant, G. A. and Bond, J. S. (1992) *J. Biol. Chem.* **267**, 9185-9193
- 6 Dummermuth, E., Sterchi, E. E., Jiang, W., Wolz, R. I., Bond, J. S., Flannery, A. V. and Benyon, R. J. (1991) *J. Biol. Chem.* **266**, 21381-21385
- 7 Kaushal, G. P., Walker, P. D. and Shah, S. V. (1994) *J. Cell. Biol.* **126**, 577-585
- 8 Chantry, A., Gregson, N. and Glynn, P. (1989) *J. Biol. Chem.* **264**, 21603-21607
- 9 Chantry, A. and Glynn, P. (1990) *Biochem. J.* **268**, 245-248
- 10 Chantry, A., Earl, C., Groome, N. and Glynn, P. (1988) *J. Neurochem.* **50**, 688-694
- 11 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn., Cold Spring Harbor Press, Cold Spring Harbor
- 12 Jackson, A. P., Seow, H.-F., Holmes, N., Drickamer, K. and Parham, P. (1987) *Nature* (London) **326**, 154-159
- 13 Sheets, M. D., Ogg, S. C. and Wickens, M. P. (1990) *Nucleic Acids Res.* **18**, 5799-5805
- 14 Kozak, M. (1986) *Cell* **44**, 283-292
- 15 Paine, M. J. I., Desmond, H. P., Theakston, R. D. G. and Crampton, J. M. (1992) *J. Biol. Chem.* **267**, 22869-22876
- 16 Kini, R. M. and Evans, H. J. (1992) *Toxicol.* **30**, 265-293
- 17 Van Wart, H. E. and Birkedal-Hansen, H. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5578-5582
- 18 Grams, F., Huber, R., Kress, L. F., Moroder, L. and Bode, W. (1993) *FEBS Lett.* **335**, 76-80
- 19 Smekens, S. P. (1993) *Biotechnology* **11**, 182-186
- 20 Halban, P. A. and Irminger, J.-C. (1994) *Biochem. J.* **299**, 1-18
- 21 Pei, D. and Weiss, S. J. (1995) *Nature* (London) **375**, 244-247
- 22 Sato, H., Takino, T., Okada, Y., Coa, J., Shinagawa, A., Yamamoto, E. and Seiki, M. (1994) *Nature* (London) **370**, 61-65
- 23 Bjarnason, J. B. and Fox, J. W. (1995) *Methods Enzymol.* **248**, 345-368
- 24 Blobel, C. P., Wollberg, T. G., Turck, C. W., Myles, D. G., Primakoff, P. and White, J. M. (1992) *Nature* (London) **356**, 248-252
- 25 Perry, A. C. F., Jones, R., Barker, P. J. and Hall, L. (1992) *Biochem. J.* **286**, 671-675
- 26 Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1990) *J. Cell Biol.* **111**, 69-78
- 27 Wollberg, T. G., Bazan, J. F., Blobel, C. P., Myles, D. G., Primakoff, P. and White, J. M. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10783-10787
- 28 Yagami-Hirosama, T., Sato, T., Kurisaki, T., Kamijo, K., Nabeshima, Y. and Fujisawa-Sehara, A. (1995) *Nature* (London) **377**, 652-656
- 29 Yoshida, S., Setoguchi, M., Higuchi, Y., Akizuki, S. and Yamamoto, S. (1990) *Int. Immunol.* **2**, 585-591
- 30 Katagiri, T., Harada, Y., Emi, M. and Nakamura, Y. (1995) *Cytogenet. Cell Genet.* **68**, 39-44
- 31 Emi, M., Katagiri, T., Harada, T., Saito, H., Inazawa, J., Ito, I., Kasumi, F. and Nakamura, Y. (1993) *Nature Genet.* **5**, 151-157
- 32 Jiang, W. and Bond, J. S. (1992) *FEBS Lett.* **312**, 110-114
- 33 Bode, W., Gomis-Ruth, F.-X. and Stockler, W. (1993) *FEBS Lett.* **331**, 134-140
- 34 Rawlings, N. D. and Barrett, A. J. (1995) *Methods Enzymol.* **248**, 183-228